Lesson 5: Solving problems with ML models

Experience AI



Order the stages of the AI project lifecycle



Order the stages of the AI project lifecycle





Lesson 5: Solving problems with ML models

In this lesson, you will:

- Describe the stages of the AI project lifecycle
- Use a machine learning tool to import data and train a model
- Test and examine the accuracy of a machine learning model

User-focused

A 'user-focused' approach to creating Al applications means focusing on the problems faced by users and taking into account their wants and needs.

Al applications have the potential to disadvantage or even pose dangers to humanity if used inappropriately.

A user-focused approach aims to avoid harm by focusing on benefitting users over making profit.

Question:

Can you suggest one way in which you could try to avoid the use of an Al application causing harm?

United Nations Sustainable Development Goals

The United Nations have defined 17 goals that act as a framework for building a better future for everyone.

Are there ways in which machine learning models can be created to **reduce inequalities** in the world?



Image source:

https://www.un.org/sustainabledevelopment/news/communications-material

United Nations Sustainable Development Goals



Image source: https://www.un.org/sustainabledevelopment/news/communications-material

Activity 1

Solving problems using ML models

Over the next two lessons, you will be tasked with creating a **classification model** to solve a real-world problem.

To do this, you will work through the stages of the AI project lifecycle.

Choose a project



Classifying data from the world's oceans



Identifying fake news



Classifying organic and non-organic waste

Spend 5 minutes looking at the project briefs and choose the problem to solve that interests you the most.

Activity 1



Watch the video on YouTube

Stage 1: Defining the problem

The problem has been outlined and you have been provided with the data to be used.

Your task is to:

- Consider who will benefit from the solution
- Justify why a data-driven approach is suitable for solving this problem



At this stage of the AI project lifecycle, the problem that needs to be addressed is clearly outlined.

Also outlined and agreed is:

- Who the intended users are and the benefits and risks to creating the solution
- Why creating an ML model is a suitable approach
- What data will be collected and from where

Activity 1

Stage 2: Preparing the data

A machine learning model will only ever be as successful as the data that is used to train it.

Deciding on what data to use is an important part of this, but equally as important is '**cleaning**' the data you have.

Can you spot any issues with this data?

Name	Animal type	Weight (kg)	Height (m)	Age	Gender
Echo	Elephant	5900	3.3	45	Female
Stretch	Giraffe	800	5.9		Male
Yakov	Meerkat	0.73	0.32	11	Male
Maiya	Meerkat	0.67	0.33	11	Female
Vassily	Meerkat	0.69	0.32	4	Male
Bogdan	Meerkat	0.76	3	1	Male
Spot	Giraffe	820	5.1	1	Male
Batyr	Elephant	6000	3.2	32	Male
Sher	Lion		1.1	7	Female
Lavi	Lion	130	1.2	4	Female
Sarabi	Lion	124	1.3	4	Female
Drona	Elephant	0	3.1	9	Male
Alexander	Meerkat	0.71	3	4	Male
Scar	Lion	190	1.3	8	Male
Beo	Giraffe	750	4.9	5	Female
Alexander	Meerkat	0.71	3	4	Male

Missing values

Name	Animal type	Weight (kg)	Height (m)	Age	Gender
Echo	Elephant	5900	3.3	45	Female
Stretch	Giraffe	800	5.9		Male
Yakov	Meerkat	0.73	0.32	11	Male
Maiya	Meerkat	0.67	0.33	11	Female
Vassily	Meerkat	0.69	0.32	4	Male
Bogdan	Meerkat	0.76	3	1	Male
Spot	Giraffe	820	5.1	1	Male
Batyr	Elephant	6000	3.2	32	Male
Sher	Lion		1.1	7	Female
Lavi	Lion	130	1.2	4	Female
Sarabi	Lion	124	1.3	4	Female
Drona	Elephant	0	3.1	9	Male
Alexander	Meerkat	0.71	3	4	Male
Scar	Lion	190	1.3	8	Male
Beo	Giraffe	750	4.9	5	Female
Alexander	Meerkat	0.71	3	4	Male

Activity 1

Duplicate entry

Name	Animal type	Weight (kg)	Height (m)	Age	Gender
Echo	Elephant	5900	3.3	45	Female
Stretch	Giraffe	800	5.9		Male
Yakov	Meerkat	0.73	0.32	11	Male
Maiya	Meerkat	0.67	0.33	11	Female
Vassily	Meerkat	0.69	0.32	4	Male
Bogdan	Meerkat	0.76	3	1	Male
Spot	Giraffe	820	5.1	1	Male
Batyr	Elephant	6000	3.2	32	Male
Sher	Lion		1.1	7	Female
Lavi	Lion	130	1.2	4	Female
Sarabi	Lion	124	1.3	4	Female
Drona	Elephant	0	3.1	9	Male
Alexander	Meerkat	0.71	3	4	Male
Scar	Lion	190	1.3	8	Male
Beo	Giraffe	750	4.9	5	Female
Alexander	Meerkat	0.71	3	4	Male

Invalid data (outside normal range)

Name	Animal type	Weight (kg)	Height (m)	Age	Gender
Echo	Elephant	5900	3.3	45	Female
Stretch	Giraffe	800	5.9		Male
Yakov	Meerkat	0.73	0.32	11	Male
Maiya	Meerkat	0.67	0.33	11	Female
Vassily	Meerkat	0.69	0.32	4	Male
Bogdan	Meerkat	0.76	3	1	Male
Spot	Giraffe	820	5.1	1	Male
Batyr	Elephant	6000	3.2	32	Male
Sher	Lion		1.1	7	Female
Lavi	Lion	130	1.2	4	Female
Sarabi	Lion	124	1.3	4	Female
Drona	Elephant	0	3.1	9	Male
Alexander	Meerkat	0.71	3	4	Male
Scar	Lion	190	1.3	8	Male
Beo	Giraffe	750	4.9	5	Female
Alexander	Meerkat	0.71	3	4	Male

Stage 2: Preparing the data

The data provided for you has already been cleaned and is ready to be used to train a machine learning model.

As you will be using **classification** to solve this problem, you must plan what classes you will use to label the data.



Activity 1



Stage 3: Training a machine learning model

You have the data, and you have planned your classes.

The next stage is to train the model.

Before you do this, consider how much data you will use to **train** the model and how much you need to set aside to **test** the model later. Factors to consider:

- Using too little **training data** can mean the model is likely to make less accurate predictions
- Using too much training data means you have less test data to use to help understand how accurate the model's predictions are

Activity 2

Demonstration of training a model

Stage 4: Testing the model

Earlier you set aside some of the data to be used for testing.

Your next task is to test the model with some of this data and to measure the **accuracy**.



Measuring accuracy during testing

There is a simple calculation you can do to measure the **accuracy** of your model:

Accuracy = number of correct predictions total number of predictions

This number can be multiplied by 100 to give the answer as a percentage.

Worked example:

- 13 tests were carried out
- 11 predictions were correct
- 11 ÷ 13 = 0.84

- 0.84 × 100
- Accuracy = 84%

Accuracy and confidence

ML classification models do not just return a label, they also return a **confidence score**.

A minimum level of confidence (**threshold**) should be decided on.

A model might return the correct label, but we might not be satisfied with the level of confidence.



Recognised as **Apples** with 58% confidence

Accuracy and confidence example

What would you set as the confidence threshold for an ML model used in an application to help **predict the weather in your area**?

- Ideally 100%?
- Would 80% be acceptable?



Accuracy and confidence example

What would you set as the confidence threshold for an ML model used by **a driverless car to identify pedestrians crossing the road**?

- Ideally 100%?
- Would 20% be better?



Measuring accuracy with a confidence threshold

Accuracy is now the proportion of predictions where:

• The label is correct

and

• The confidence threshold is met

Worked example:

- 13 tests were carried out
- 11 predictions were correct, but the confidence threshold was met for only 8 of these
- 8 ÷ 13 = 0.62
- 0.62 × 100
- Accuracy = 62%

Test your model

Go to **Stage 4** on your worksheet.

You must:

- Decide on a confidence threshold
- Test your model with at least 10 examples of test data
- Work out the accuracy of your model
- Reflect on the results



Reporting on the accuracy of a model

Student A

Accuracy of model = 56%

Student **B**

Accuracy of model = 95%

Both students have trained a model aimed at **identifying fake news**.

- Which model would you use?
- What information is missing to help you make this decision?

Reporting on the accuracy of a model

Student A

Accuracy of model = 56%

Confidence threshold = 100%

Student B

Accuracy of model = 95%

Confidence threshold = 60%

Both students have trained a model aimed at **identifying fake news**.

- Which model would you use?
- What information is missing to help you make this decision?

Summary

Next lesson

In this lesson, you...

Worked through stages of the AI project lifecycle

Trained a machine learning model

Tested and examined the accuracy of the model

Next lesson, you will...

Evaluate and explain your models

Explore AI-related careers